chapter : 7 (Interference concerning proportion)

7.1) Estimation of proportion.

The procedure of testing hypothesis is,

i) set up the null hypothesis H_0

ii) set up the alternate hypothesis H_1

iii) Decide appropriate level of significance.

~~iv) compute null hypothesis~~

iv) compute the test statistic under null hypothesis

$$z = \frac{\bar{p} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$$

v) we reject H_0 if $z \geq z_\alpha$ for one tailed test

$z \geq z_{\alpha/2}$ for two tailed test

vi) Draw a conclusion about accept or reject Null hypothesis

7.2) Hypothesis concerning one proportion.

Procedure for testing hypothesis are

i) set up null hypothesis $H_0 : P = P_0$

ii) set up alternate hypothesis $H_1 : P \neq P_1$ (for two tailed test)

$H_1 : P > P_1$ (for right tailed test)

$H_1 : P < P_1$ (for left tailed test)

iii) level of significance (α)

iv) Test statistic $(z) = \dfrac{\bar{p} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$

v) obtain critical value for significant α

vi) we reject H_0 if $|z| \geq z_\alpha$ for one tailed test

$|z| \geq z_{\alpha/2}$ for two tailed test

vii) Draw a conclusion about accept or reject null hypothesis

chapter: 7

7.3) Hypothesis concerning two proportion

Procedure for testing hypothesis are,

i) set up null hypothesis $H_0 : P_1 = P_2$

ii) set up alternate hypothesis $H_1 : P \neq P_1$ (two tailed test)

$H_1 : P > P_1$ (right tailed test)

$H_1 : P < P_1$ (left tailed test)

iii) level of significance (α)

iv) Test statistics $(z) = \dfrac{(\bar{P}_1 - \bar{P}_2) - (P_1 - P_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

If $\bar{p}$ is not given then, $\bar{p} = \dfrac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2}$

v) obtain critical value for significant α

vi) we reject H_0 if $|z| \geq z_\alpha$ for one tailed test

$|z| \geq z_{\alpha/2}$ for two tailed test

vii) Draw a conclusion about accept or reject null hypothesis

070 chaitra

Q) A manufacturer of submersible pumps claims that at most 30% of the pumps require ~~to~~ repair within the first 5 years of operation. If a random sample of 120 of these pumps includes 47 which require repair within the first 5 years, test the null hypothesis $P = 0.30$ against the alternate hypothesis $P > 0.30$ at the ~~level of signif~~ 0.05 level of significance.

Soln Given, sample size $(n) = 120$

no. of pumps which requires repair $(x) = 47$

so, $\bar{p} = \frac{x}{n} = \frac{47}{120} = 0.39167$

We set up the following hypothesis

1) Null hypothesis $H_0 : P = P_0 = 0.30$

ii) Alternate hypothesis $H_1 : P > 0.30$

iii) level of significance, $\alpha = 0.05$

iv) Test statistic

$$z = \frac{\bar{p} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$$

criteria

v) we reject H_0 if $|z| \geq z_{0.05} = 1.645$

vi) calculation & decision

$$\bar{p} = \frac{x}{n} = \frac{47}{120} = 0.39167$$

$$\therefore z = \frac{\bar{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.39167 - 0.3}{\sqrt{\frac{0.3 \times 0.7}{120}}}$$

$$= 2.19$$

since $z = 2.19 > 1.645$ (tabular value)

we reject H_0 & conclude that more than 30% pumps require repair within 5 years of operation.

Q) Define confidence level & ~~le~~ significance level. 16 1

chapter: 7

07/shrawn 13 Q) Define confidence level & significance level. A manufacturer claimed that at least 95% of the cables supplied to the ABC company confirmed to specifications. However, the production manager at ABC company wasn't satisfied with the claim of the manufacturer. Hence to test the claim, the manager examined a sample of 250 cables supplied last month and found that 228 cables work as per specification. can you conclude that the production manager is right to doubt on the claim of the manufacture? ($\alpha = 0.01$).

⇒ The risk of type I error which is tolerated in making a decision about H_0 is known as level of significance. It is denoted by α

And if α is subtracted by 1 ~~the~~ ie $(1-\alpha)\%$ then it is confidence level.

Soln Given, sample size(n) = 250

no. of cable as per specification (x) = 228

so, $\bar{p} = \frac{x}{n} = \frac{228}{250} = 0.912$

We set up the following hypothesis,

i) Null hypothesis

$H_0: P = P_0 = 0.95$ (production manager is not right to doubt on claim)

ii) Alternate hypothesis

$H_1: P < 0.95$ (production manager is right to doubt on claim)

iii) <u>level of significance</u>

$\alpha = 0.01$

iv) <u>Test statistics</u>

$$z = \frac{\bar{p} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$$

v) <u>criteria</u>

We reject H_0 if $|z| \geq Z_\alpha = Z_{0.01} = \cancel{2.326}\ 2.576$

vi) <u>calculation & decision</u>

$\bar{p} = 0.912$

$P_0 = 0.95$

$q_0 = 1 - P_0 = 1 - 0.95 = 0.05$

$n = 250$

Now, $z = \dfrac{\bar{p} - P_0}{\sqrt{P_0 q_0 / n}} = \dfrac{0.912 - 0.95}{\sqrt{\frac{1}{250} \times 0.95 \times 0.05}} = -2.757$

since $|z| \geq Z_\alpha$ i.e $\cancel{2.326}\ 2.576$

so, we reject H_0 & conclude that production manager was right to doubt on the claim.

1:10 Ahad 148) Two different types of injection-molding machines are used to form plastic parts. A part is considered defective if it has excessive shrinkage or is discolored. Two random samples, each of size 300, are selected & 15 defective parts are found in the sample from machine 1 while 8 defective parts are found in the sample from machine 2. It is resonable to conclude that both machines produce the same fraction of defective parts, using $\alpha = 0.05$?

Sol^n We set up the following hypothesis

i) Null hypothesis

$H_0: P_1 = P_2$ (both machine produce same fraction of defective parts)

ii) Alternate hypothesis:

$H_1: P_1 \neq P_2$ (either one machine produce more fraction of defective parts).

iii) level of significance

$\alpha = 0.05$

iv) Test statistics

$$Z = \frac{\bar{P}_1 - \bar{P}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where, $\bar{p} = \frac{n_1\bar{P}_1 + n_2\bar{P}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$

v) criteria

We reject H_0 if $|z| \geq z_\alpha = z_{0.05} = 1.645$

vi) Calculation & decision

$$\bar{P}_1 = \frac{x_1}{n_1} = \frac{15}{300} = 0.05$$

$$\bar{P}_2 = \frac{x_2}{n_2} = \frac{8}{300} = 0.0267$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{15+8}{300+300} = 0.0383$$

$\bar{q} = 1 - \bar{p}$

$$\therefore Z = \frac{0.05 - 0.0267}{\sqrt{0.0383 \times 0.9617\left(\frac{1}{300} + \frac{1}{300}\right)}} = 1.4869$$

Since $|z| \ngtr z_\alpha = 1.645$ ~~we so~~ we donot reject H_0 & conclude that both machine produce same fraction of defective parts.

069 chaitra

Q) What are the steps in hypothesis testing? A study shows that 16 of 200 computers produced on one assembly need readjustment before shipping while same happens on 14 out of 300 produced. Test at 1% level of significance that the second assembly is superior than first one?

⇒ The steps in hypothesis testing are given below.

i) set up the null hypothesis H_0

ii) set up the alternate hypothesis H_1. It may be either one tailed or two tailed.

iii) Decide appropriate level of significance.

iv) compute the test statistics under null hypothesis

$$z = \frac{t - E(t)}{S.E(t)}$$

v) Draw a conclusion about accept or reject null hypothesis H_0 with the help of comparision between value of test statistic z & critical values $z_{\alpha/2}$ (for two tailed test) or z_α (for one tailed test) at α level of significance

a) For two tailed test

If $|z| > z_{\alpha/2}$ we reject null hypothesis H_0

b) For one-tailed test

If $|z| > z_\alpha$ we reject null hypothesis otherwise we accept H_0. If we accept our null hypothesis, we say, there is no significant difference between statistic & population parameter. This implies that the difference $t - E(t)$ is just due to fluctuation of sampling.

Soln We set up the following hypothesis.

i) Null hypothesis

$H_0: P_1 = P_2$ (there is no significance difference)

ii) Alternative hypothesis

$H_1: P_2 > P_1$ (second assembly is superior to first one).

iii) level of significance

$\alpha = 0.01$

iv) Test statistics

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{p} = \dfrac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2}$

v) criteria

we reject H_0 if $|z| > z_\alpha = z_{0.01} =$ ~~2.326~~ 2.576

vi) calculation & decision

$$\bar{p}_1 = \frac{x_1}{n_1} = \frac{16}{200} = 0.08$$

$$\bar{p}_2 = \frac{x_2}{n_2} = \frac{14}{300} = 0.0467$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{16+14}{200+300} = 0.06$$

$$\& \ \bar{q} = 1 - \bar{p} = 1 - 0.06 = 0.94$$

$$\therefore \ Z = \frac{\bar{P}_1 - \bar{P}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.08 - 0.0467}{\sqrt{0.06 \times 0.94 \left(\frac{1}{200} + \frac{1}{300}\right)}}$$

$$= 1.5360$$

since $|z| \ngtr Z_\alpha$ i.e. ~~2.326~~ 2.576 so we accept H_0 & conclude that there is no significant difference between first & second assembly.

08 chaitra 14Q) A soft drink is being bottled using two different filling machines The standard deviation of the process for machine A & B was 0.010 & 0.015L respectively. 30 bottles were randomly sampled from each machine & the mean were 2.04 & 2.07L respectively. Can one conclude that both machines are filling the same volume of soft drink? Test the hypothesis at $\alpha = 0.01$ level of significance.

solⁿ we set up the following hypothesis

i) Null hypothesis

$H_0: \mu_1 = \mu_2$ (both machine are filling same volume)

ii) Alternative hypothesis

$H_1: \mu_1 \neq \mu_2$ (both machine are not filling the same volume).

iii) level of significance

$\alpha = 0.01$

iv) Test statistics

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ for $n_1 \geq 30$ & $n_2 \geq 30$

v) criteria

we reject H_0 if $|z| > Z_\alpha = 2.576$

vi) calculation & conclusion

Given,

$n_1 = 30$, $\bar{x}_1 = 2.04$, $s_1 = 0.010$

$n_2 = 30$, $\bar{x}_2 = 2.07$, $s_2 = 0.015$

so,

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.04 - 2.07}{\sqrt{\frac{(0.01)^2}{30} + \frac{(0.015)^2}{30}}}$$

$= -9.1146$

Since $|z| > Z_\alpha$ i.e 2.576 so we ~~accept~~ reject H_0 & conclude that both machine are not filling same volume.

7.4) chi-square test of independance.

χ^2 test is a non parametic test because it depends only on the set of observed & expected frequencies & degree of freedom. since χ^2 test doesnot works any assumption about population parameter x, it is also called a distribution free test.

chapter: 7

It is defined as,

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

where, O = observed frequency
E = expected frequency

- procedure for χ^2 test.

1) <u>Null hypothesis H_0</u>:
There is no significant difference between observed & expected frequencies.

11) <u>Alternative hypothesis H_1</u>:
There is significant difference between observed & expected frequencies.

111) <u>Test statistics</u>

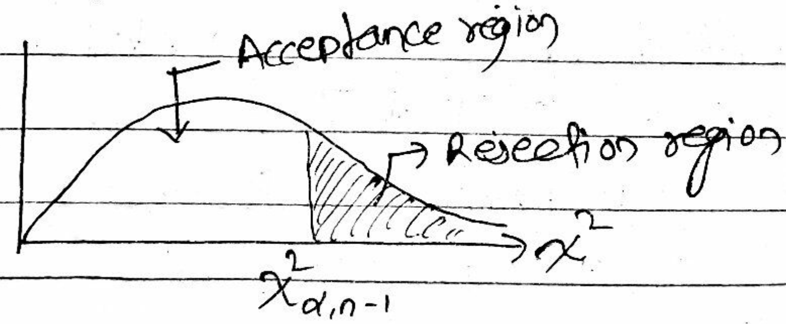
$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

1v) Find critical value from table in reference to level of significance α (usually 1% & 5%)

v) <u>Decision</u>

a) If calculated value of χ^2 > tabulated value of χ^2 then we reject H_0 & accept H_1.

b) If calculated value of $\chi^2 \leq$ tabulated value of χ^2 then we accept H_0 & reject H_1.



If no significant difference then, poisson distribution can be followed

8) Define chi-square distribution

12·3

8) Define chi-square distribution. A book containing 500 pages was throughly checked. The distribution of number of error page was given below as.

No. of errors	0	1	2	3	4	5
No. of pages	275	138	75	7	4	1

Use chi-square test of goodness of fit, verify whether the arrivals follows a poisson distribution at 5% level of significance

⇒ chi-square distribution is defined as the sum of ratio of square of difference between observed & expected frequency to that expected frequency. And is given as

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

where, O = observed frequency &
E = Expected frequency

sol? Here,

	No. of errors						
	0	1	2	3	4	5	Total
No. of pages	275	138	75	7	4	1	500
	275	138	75	7	4	1	500

d.f = ~~(r-1)(c-1) = 1.(6-1) = 5~~
(n-1) = 6-1 = 5

We set up the following hypothesis

1) Null hypothesis H_0:
There is no significant difference between observed & expected frequencies.

ii) Alternative hypothesis
There is significant difference between ~~two~~ observed & expected frequencies.

iii) level of significant

$\alpha = 5\% = 0.05$

iv) Test statistic

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

v) criteria

We reject H_0 if $\chi^2 > \chi^2_{0.05,5} = 11.070$

vi) observation & calculation

No. of errors	No. of pages (O)	No. of pages (E) = Total/6	$(O-E)^2/E$
0	275	~~100~~ 83.33	440.87
1	138	~~100~~ 83.33	35.87
2	75	~~100~~ 83.33	0.83
3	7	~~100~~ 83.33	69.92
4	4	~~100~~ 83.33	75.52
5	1	83.33	81.34
			704.35

so, $\chi^2 = \Sigma \frac{(O-E)^2}{E} = 704.35$.

Since $\chi^2 > \chi^2_{0.05,5} = 11.070$ (tabulated)

so, we reject Ho & conclude that there is significant difference between expected & observed frequencies & the arrival doesnot follow a poisson distribution.

072 chaitra)

13 Q) Define chi-square distribution. From the following data can you conclude that there is association between the purchase of brand & geographical region? ($\alpha = 5\%$)

	Region		
	central	Eastern	western
purchase brand	40	55	45
do not purchase brand	60	45	55

⇒ Sol^n Here.

	Region			Total
	central	Eastern	western	
purchase brand	40	55	45	140
do not purchase brand	60	45	55	160
	100	100	100	300

degree of freedom, $d.f = (r-1)(c-1)$

$= (2-1)(3-1) = 1 \times 2 = 2$

Here, $E_{11} = E(40) = \frac{100 \times 140}{300} = 46.67$

$E_{12} = E(55) = \frac{100 \times 140}{300} = 46.67$

$E_{13} = E(45) = \frac{100 \times 140}{300} = 46.67$

$E_{21} = E(60) = \frac{100 \times 160}{300} = 53.33$

$E_{22} = E(45) = \frac{100 \times 160}{300} = 53.33$

$E_{23} = E(55) = \frac{100 \times 160}{300} = 53.33$

observed frequency O_{ij}	expected frequency E_{ij}	$\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$
40	46.67	0.95
55	46.67	1.49
45	46.67	0.06
60	53.33	0.83
45	53.33	1.30
55	53.33	0.05
		4.68

We set up the following hypothesis.

i) Null hypothesis H_0:
There is no significant difference between observed & expected frequency

ii) Alternative hypothesis H_1:
There is significant difference between observed & expected frequency.

iii) level of significant
$\alpha = 5\% = 0.05$

iv) Test statistic
$$\chi^2 = \Sigma \frac{(O-E)^2}{E} \cancel{= 4.68}$$

v) criteria
We reject H_0 if $\chi^2_{cal} > \chi^2_{0.05,2} = 5.992$

vi) observation & calculation
$$\chi^2_{cal} = \Sigma \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = 4.68$$

since $\chi_{cal} \ngtr \chi^2_{0.05,2}$ i.e 5.992 so we accept Ho & conclude that there is association between the purchase of brand & geographical region.